Abstract

An algorithm has been developed to identify four DNA sequences of 20 bases or more that form a structure called a connectron. Two sequences C1 and C2 are adjacent to each other. These sequences are expressed as RNA in the 3'UTR of some genes in many prokaryotic, archea and eukaryotic genomes. The other half of a connectron is two DNA sequences T1 and T2 that are on the same chromosome and range in distance from each other by about 1kb to 105kb. The C1 sequence is identical to the T1 sequence and the C2 sequence is identical to the T2 sequence. C1/C2 and T1-T2 can be on different chromosomes. The C1/C2 RNA sequence of the gene transcript finds the two double-stranded DNA sequences T1 and T2. The single-stranded RNA and double-stranded DNA then form a triple-stranded Hoogsteen helix of the RNA/DNA/DNA variety. Because the C1 sequence is adjacent to the C2 sequence, the T1 sequence is made spatially adjacent to the T2 sequence in a compact X-shaped structure. Chromatin particles form as compact 30nm assemblies in the DNA between T1 and T2 thus eliminating the intervening genes from promotion and expression. Connectrons remove sets of genes from expression and thus modulate the behavior of many types of cells.